Coding challenge: Which products will be successful?

Task Summary

A fashion e-commerce company is planning its collections for the upcoming year. Therefore the company put together many potential products as candidates and now would like to estimate which products would be successful (top) or not (flop). To do so, you are provided with data of the past years' top and flop products. This will allow you to create a small machine learning application (without an UI).

Given Data

You have two data sets:

- Historic data: Products of the past two years and their attributes (including a label that categories the item as top or flop); file: historic.csv (8000 products)
- Prediction input data: Potential products of the upcoming year and their attributes (but no label about the success); file: prediction_input.csv (2000 product candidates)

Extract of the historic data:

item_no	category	main_promotion	color	stars	success_indicator
737192	Hoodie	Category_Highlight	White	4.9	top
240528	T-Shirt	Display_Ad_Campaign	Yellow	3.9	top
775333	T-Shirt	Catalog	Black	3.1	flop
880358	Blouse	Category_Highlight	Blue	2.8	flop
237635	Hoodie	Catalog	Green	3.8	top

Columns:

- item_no: Internal identifier for a past product or a product candidate for the future.
- category: Category of the product.
- main_promotion: Main promotion that would be / was used to promote the product.
- color: The main color of the product.
- stars: Stars of reviews from a comparable product of a competitor (from 0 = very negative reviews to 5 = very positive reviews).
- success_indicator: Indicator whether a product was successful (top) or not (flop) in the past. Only given for the historic data.

Technical Requirements

- We would like to see a small object oriented, class based python ML application to perform preprocessing, training and prediction (with either an ANN or a decision tree based model) without any UI.
- The specific accuracy of your model is of minor importance.
- Please follow common conventions.
- Jupyter notebooks shall not be used for the ML application but can be used for descriptive analysis (if helpful).
- Please use (short) docstrings to document your code.
- Please use Python 3.7+ and current versions of common libraries.

Notes

- You will agree to only submit your own work! You are free to do research online or in relevant literature.
- The given data sets are stochastically generated and are not real data of our company. Nevertheless it is possible to make proper predictions based on the attributes.
- The task can be completed in less than 9 hours. Please submit within 3 days after obtaining the challenge.